REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
04/21/2014	Final	6/1/2011 - 12/31/2013

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Towards Optimization of Macrocognitive Processes: Automating Analysis of the Emergence of Leadership in Ad Hoc Teams	5b. GRANT NUMBER N000141110221
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Rosé, Carolyn, Penstein; Stahl, Gerry; Goggins, Sean; Patterson, Emily; Borge, Marcela; Carroll, John; Duchon, Andrew	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213-3890, Drexel University (Philadelphia, PA), Ohio State University (Columbus, OH), Penn State University (University Park, PA), Aptima, Inc. (Woburn, MA)	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Office of Naval Research 875 North Randolph Street Arlington, VA 22203-1995	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

13. SUPPLEMENTARY NOTES

14. ABSTRACT

An important focus for technical research related to machine learning has been to address not only generalization across sub-community structures within a hierarchical dataset, but also accommodating changes over time in a longitudinal dataset using evolving behavior models. We have two prototype models built and working and are extending that work to make it more scalable to larger datasets. We completed a highly scalable model, being able to be applied to networks with millions of users. We validated the model on 3 different datasets from Massive Open Online Courses and found that the sub-community structure identified by the algorithm was predictive of differences in dropout rate between subsets of students.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Carolyn Penstein Rosé
					19b. TELEPHONE NUMBER *(Include area code)*
				25	412.268.7130

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Final Report for ONR N000141110221 – April 2014

Title of Proposal: Towards Optimization of Macrocognitive Processes: Automating Analysis of the Emergence of Leadership in Ad Hoc Teams

Prime Offeror: (PI) Carolyn Penstein Rosé (Language Technologies Institute/ Human-Computer Interaction Institute, CMU), (CoPI) Gerry Stahl, Drexel University, (CoPI) Sean Goggins, Drexel University, (CoPI) Emily Patterson, Ohio State University, (CoPI) Marcela Borge, Penn State, (CoPI) John Carroll, Penn State, (CoPI) Andrew Duchon, Aptima

Technical Contact: Carolyn Penstein Rosé, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, phone 412-268-7130, fax 412-268-6298, cprose@cs.cmu.edu

Administrative Contact: Krista McGuigan, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, phone 412-268-6592, fax 412-268-6298, klmcg@cs.cmu.edu

Grant Start Date: June 1, 2011

SCIENTIFIC AND TECHNICAL OBJECTIVES

The objective of the Ad Hoc Teams project was to facilitate emergence of **shared leadership** in ad hoc teams through **context sensitive** support to enable proactive decision making. What this means is optimization of group knowledge construction, leading to the formation of more effective plans in less time. We can only meet this objective if we facilitate getting the right information to the right people at the right time **AND** getting the right people to contribute their expertise at the right time. The key to enabling the basic research to understand how to design such support as well as the technology to provide the support in real time is a foundation in machine learning and language technologies, which underlie an infrastructure that speeds up the data to actionable knowledge loop. A key component of this effort is the establishment of a central repository for CDM data and analyses, the Combined Canonical CDM Corpus (C4), which will play a central role in the data to actionable knowledge loop, and will provide a valuable **resource for the whole CDM community.** Seven years of successful evaluation studies of the basic architecture for technology supported collaboration developed in our prior work provides a strong demonstration of its potential impact on task success for group knowledge construction and strategic planning tasks. Conversation technology offers interactive support for teams.

APPROACH

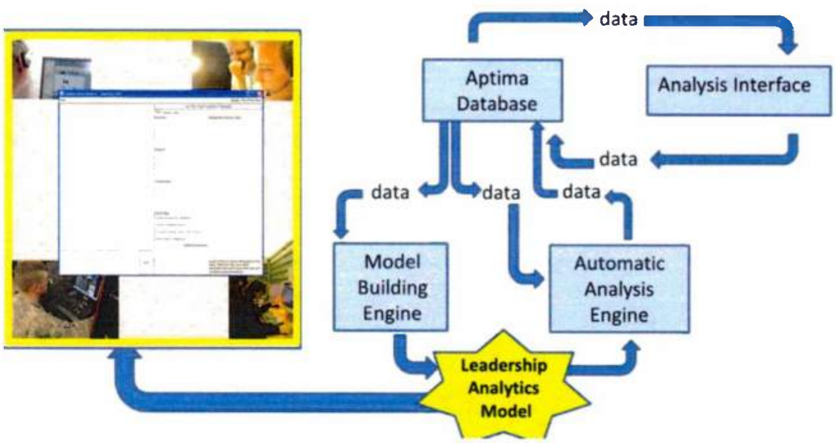


Figure 1 Graphical representation of our reusable architecture for supporting the data to actionable knowledge loop, which embodies our technical approach. On the left we see an example interface for supporting distributed collaboration in the Non-combatant Evacuation Scenario designed as part of the CKI program under Norman Warner. Context sensitive support for this task was triggered based on real time monitoring of the collaboration from four participants working synchronously, but not co-located. The automated analysis engine (bottom right) was trained using annotated data from an earlier study (see analysis infrastructure in upper right).

Our end goal is to use ***Leadership analytics models*** trained from annotated data using machine learning to ***monitor collaboration in real time*** and ***trigger context sensitive support*** that will

increase mission success in measurable ways. In order to make this happen, we are pursuing a mid-level goal to improve CDM program infrastructure to ***increase the efficiency of the program Data to Knowledge loop*** by supporting both ***human analysis*** and ***automatic analysis.*** A key component of this technical approach is the development of a central repository of CDM datasets, integrated with tools to support human analysis as well as text mining tools to support automated analysis.

CONCISE ACCOMPLISHMENTS

In the final year of the project, as in the years leading up to that, we have produced both scientific accomplishments as well as practical ones.

On a practical level, throughout the project we consistently engaged in substantial development. By the end, we completed the technical machinery needed to enable the data to actionable knowledge loop that has been a key aspect of our proposed work. This effort has been lead by Co-PIs Goggins and Duchon. At the June PI meeting, we were able to demonstrate the full pipeline. This infrastructure is designed to work in real-time, such that analyses that are developed could be applied more easily to real-world domains, both civilian and military. In the final months of the project, we continued to collect, process, and insert datasets in to the central repository that is a key component in this pipeline.

In terms of scientific accomplishments in the area of *Machine Learning*, PI Rosé has lead the effort to advance technology for machine learning that enables taking advantage of the domain structure and subpopulation structure of a hierarchically structured corpus (as are all corpora with group interaction data due to interaction between people within groups that introduces dependencies between the behaviors of individuals within those groups) in achieving high classification performance.

Pairing these two accomplishments, we are now able to more easily apply the analytic tools we have developed to producing new knowledge in the area of group science by facilitating analyses of new data. At the June PI meeting, we presented results comparing the capabilities of the automated leadership analysis produced collaboratively by Co-PIs Duchon and Patterson on a dataset collected by Co-PI Borge. As a validation, we compared the automated analysis to a hand analysis done previously by Co-PI Borge. The comparison was interesting both in terms of revealing how accurate overall the automated analysis was, but what interesting limitations were identified in an error analysis that suggest important directions for our continued modeling work.

EXPANDED ACCOMPLISHMENTS

The official start date of the project was June 1, 2011. We submitted a Whitepaper with our revised plans responding to feedback from the Program Review meeting in August 2011. The full set of technical objectives in that white paper included:

1. Build robust, integrated technical infrastructure for conducting automated analysis of multiple existing, coded datasets
    a. Build common data format and combined dataset to facilitate sharing and comparing (Aptima)
    b. Develop technical infrastructure for making analysis technologies inter-operable (Aptima, CMU, Drexel)
    c. Advance machine learning technology to make it more robust and domain general (CMU)
    d. Further develop analytic techniques for identifying key positions in social networks (Drexel)
2. Develop success metrics that can be computed from interaction data and are validated against existing validated measures of macrocognition in teams
    a. Operationalization of observed success (CMU and NPS)
    b. Validation of operationalization using lab data with existing measures of macrocognition (CMU and NPS)
    c. Hand coding and automatic coding of observed success in APAN data (CMU and NPS)
3. Operationalization of emergence of shared leadership in teams
    a. Apply existing operationalizations of leadership from Penn State, Ohio State, and CMU on datasets along with Observed Success metrics and external measures of macrocognition where available (Ohio State, Penn State, and CMU)
    b. Compare relative predictive validity of alternative operationalizations (Ohio State, Penn State, and CMU)
    c. For operationalizations of leadership taking and shared leadership that predict positive outcomes, investigate the process of emergence of these processes in APAN where we can observe interactions over time (CMU)
4. Develop support for leadership taking and shared leadership in teams
    a. Automatic detection of opportunities for shared leadership (CMU and Drexel)
    b. Intelligent agents for supporting macrocognition (Ohio State and CMU)

The four proposed tasks are deeply synergistic. Task 1 provides a technological infrastructure to facilitate work on Tasks 3 and 4. Task 2 enables us to evaluate the value of operationalizations of shared leadership behaviors. Tasks 2 and 3 provide focus for continued work on Task 1, enabling the identification of key challenges faced by analysts using the technology in their basic research. In this way, we can be assured that our technological work focuses not just on what advances the fields of text mining, machine learning, and social network analysis, but that advances them in service of behavioral science that is of central importance to the CDM mission.

Tasks 1, 3 and 4 were the focus of project work in the final year and a half of the project. Below we expand upon our accomplishments in each of these three tasks leading up to the culmination of the project.

Task 1: Build robust, integrated technical infrastructure for conducting automated analysis of multiple existing, coded datasets

The purpose of Task 1 is to increase efficiency of analysis within and across datasets in order to reduce the cost of analysis while increasing rigor by facilitating more intensive triangulation across datasets. Furthermore, automatic analysis enables dynamic triggering of automatic support interventions in online collaborative environments (Kumar & Rosé, 2011; Kumar & Rosé, accepted).

Build common data format and combined dataset to facilitate sharing and comparing (Aptima)

The purpose of the C4 Database: Combined Canonical CDM Corpus has been to allow re-analysis of already gathered data. The CDM community has invested in collecting a number of corpora and has been productive in analyzing those corpora to answer the questions that motivated the data collection. However, much can be learned by comparing similarly motivated but differently specified operationalizations across corpora. Furthermore, much more can be learned from corpora as they are used as a shared resource for the CDM community to test models and methods. This supports deeper understanding of constructs, triangulation of findings, and testing of generalization across contexts. By collecting together multiple corpora that are all appropriate for investigating similar issues underlying group work and leadership taking, it is possible to leverage multiple smaller corpora as larger, heterogeneous datasets that are on a better scale to support machine learning. The technology for domain adaptation and multi-domain learning that have been a focus of this project enables us to make use of such highly heterogeneous datasets.

An important first step has been building the basic data base infrastructure that has been housed by Aptima.

Develop technical infrastructure for making analysis technologies inter-operable (Aptima, CMU, Drexel)

So far as part of the C4 Database, we have imported data from NAVAIR, OSU, PSU, ASU, APAN. The C4 Database was designed to be a universal database accommodating any kind of communication or interaction, including but not limited to Email, Chat, VoIP, Twitter, Forums, Blogs, News, Journal articles, Face to face interactions. The goal is to help reveal conceptual similarity of different data and alternative codings and allow separate development efforts to combine and share results easily.

An important part of the work we have done has been to formalize and streamline a process for cleaning up and standardizing the datasets that have been contributed. While this has been a time consuming process, it becomes more efficient with each new dataset. Co-PI Duchon's group has developed web services to aid in the standardization through an API, which means that academic groups (and other third-parties) can now access data, develop models, and apply analyses to data in real-time

Advance machine learning technology to make it more robust and domain general (CMU)

Prior work in Multi-Domain learning has assumed that a single metadata attribute (that signals a subpopulation in a dataset) is used in order to divide the data into so-called domains. However,

real-world datasets often have multiple metadata attributes that can divide the data into multi-dimensional domains. It is not always apparent which single attribute will lead to the best domains, and more than one attribute might impact classification. In our recent work (Joshi et al., 2013) we have proposed extensions to two multi-domain learning techniques for our multi-attribute setting, enabling them to simultaneously learn from several metadata attributes. Experimentally, these extensions have been demonstrated to significantly outperform the more traditional multi-domain learning baseline, even when it selects the single "best" attribute.

Further develop analytic techniques for identifying key positions in social networks (CMU and Drexel)

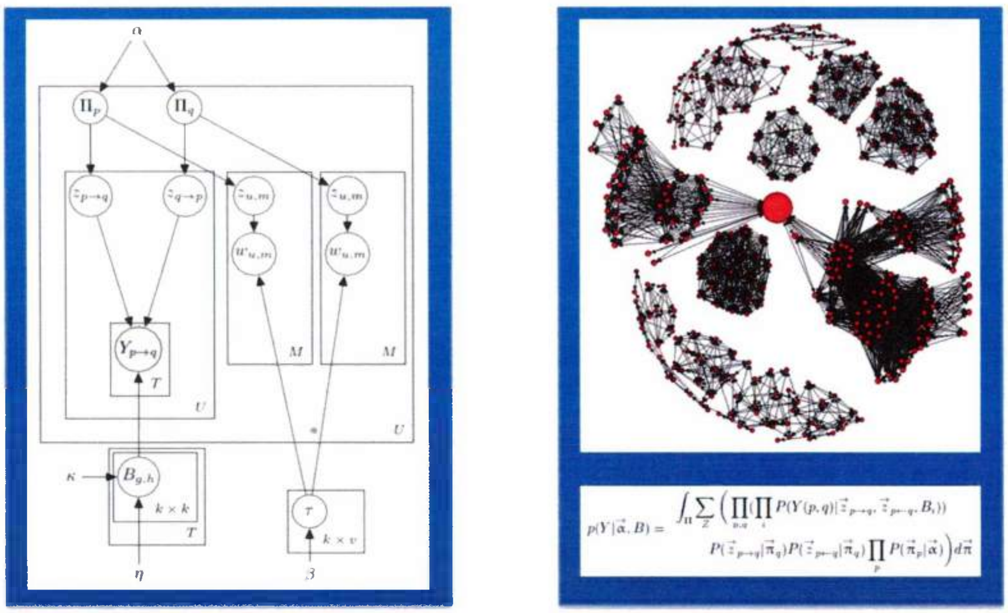


Figure 2 Tensor Based Mixed Membership Stochastic Block Models

Sometimes the subpopulation structure in large communities is latent rather than explicitly provided in meta-data features. In that case, we need to identify that structure before approaches like Mahesh's can make use of it. CMU student Abhimanyu Kumar is working on that problem using a tensor based mixed membership stochastic block model approach to graph clustering. In this more recent work, we have begun to develop a new probabilistic graphical model for modeling the dynamics of group and community level communication that leverage a similar theoretical foundation from multi-level modeling leveraged in the multi-domain learning approaches we have worked on so far but in a mostly-unsupervised setting, where generalization comes from the ability to learn structure using informative priors rather than supervision from labels.

We draw on two bodies of literature for this work. First, we draw from work on social network analysis. As a general approach, we make use of mixed membership models (Airoldi et al., 2008) that compute soft partitions of social networks (Sim et al., 2012), where each partition represents a subcommunity, and individual members can belong to and thus be influenced by the

norms associated with different ones at different times. We also draw from work integrating text mining techniques with social network analysis in order to form representations of text that reflect the community structure (McCallum et al., 2004), which builds on earlier author-topic models (Rosen-Zvi et al., 2004; Steyvers et al., 2004).

Second, we draw on recent developments in topic modeling to help identify what various communities and subcommunities are interested in. Topic modeling approaches have become very popular for modeling a variety of characteristics of unlabeled data. A well known approach is Latent Dirichlet Allocation (LDA) (Blei et al. 2003), which is a generative model and is effective for uncovering the thematic structure of a document collection. The advantage of probabilistic generative models like LDA is that it is possible to build in assumptions that bias the model in useful ways, similar to the way that structural equation models bias the estimation of weights in a linear regression based on some assumed causal structure. Two example models in prior work that are specifically tailored to the problem of modeling different perspectives are the cross-collection Latent Dirichlet Allocation (ccLDA) model (Paul and Girju, 2009) and the joint topic and perspective model for ideological discourse (Lin et al., 2008). Both assume that the frequency of a word depends on the relevance in the topic and on the perspective of the speaker or author. ccLDA (Paul and Girju, 2009) builds on the standard LDA model (Blei et al., 2003) and the cross-collection mixture model (ccMix) by Zhai et al. (2004). ccLDA discovers the topics across multiple text collections and estimates for each topic a shared distribution and collection specific distributions. The model of Lin et al. (2008) assigns every word a topical weight indicating how often it was chosen depending on the topic, and an ideological weight, which depends on the perspective of the speaker or author. In both of these cases, the subcollection structure is given to the model. What is different in our work is that the structure is found in the data based on the soft clustering described above that is based on the link structure. In that way, the community structure and the topic structure are jointly estimated and are able to influence each other.

In particular, from a technical perspective, we started with the basic mixed membership stochastic blockmodels of Airoldi (Airoldi et al., 2008). As mentioned above, the key point of a mixed membership model is that rather than each individual being assigned to one and only one community, each individual belongs probabilistically to every community. What it means that Airoldi's model is a stochastic block model is that the assumptions underlying the estimation of the model is neither as constrained as assuming a specific distribution nor as unconstrained as a non-parametric approach. As a middle ground, the distribution is assumed to be a mixture of distributions from a family, in our case the exponential family. We have made several extensions to this basic model. First, while the original model could only accommodate binary links, we were able to make the representation of connections between nodes more nuanced by enabling them to be counts or binary rather than strictly binary. Additionally, while the original model was only able to accommodate a single dimension of links, we were able to extend the model with a tensor so that it is possible to accommodate multiple dimensions of links, each representing a different perspective on relationships between nodes. Finally, we have linked the community structure that is discovered by the model with an LDA model, so that for each person a distribution of LDA topics is estimated that mirrors the distribution across subcommunities. In this way, the community structure places constraints on the topics that are estimated, and the topic structure can therefore be seen as a reflection of the community structure.

This work will allow us to structure discussion and deliberation spaces so that communities of interest can emerge, and will create the capability of connecting these communities on an ongoing basis with new documents and messages of interest to them.

Our initial extended model was limited in that it did not incorporate any notion of context (i.e., instances in time). This makes the model less interpretable than ideal. Furthermore, the initial estimation algorithm we have developed to instantiate the model from data is too computationally expensive to scale to the amount of data that we would like to apply the model to. Thus, we are currently engaged in two important directions: We are revising the structure of the model so that the association between a node, a topic, and a community will be specific to an instance in time. In this way, we can model explicitly how individuals shift over time from participation in one subcommunity and another.

At the time of submitting the last report, we had two prototype models built and working and were extending that work to make it more scalable to larger datasets. In the final months of the project, we completed such a model, which is highly scalable, being able to be applied to networks with millions of users. We validated the model on 3 different datasets from Massive Open Online Courses and found that the subcommunity structure identified by the algorithm was predictive of differences in dropout rate between subsets of students.

Because of work conducted by the Drexel team in the area of network science, the CDM program is already in a stronger position to learn valuable insights from large scale datasets like APAN. In particular, success metrics that can be computed automatically from interaction data can enable large scale automatic evaluation of developmental trajectories through online communities like APAN in terms of whether they indicate functional versus dysfunctional socialization and participation. In the time that the team has been funded, the Drexel team has already produced a high quality publication describing a network based analysis of information brokering in APAN, where information brokering can be seen as a valuable shared leadership behavior (Goggins & Mascaro, 2012).

Recent work by the Goggins group in the past fiscal year contributes to this effort. Specifically, the Goggins group has a number of publications in progress that demonstrate both tactical and strategic development of operationally relevant approaches to the identification of distributed leadership practices across a range of situations. A significant issue for war fighters is triaging and integrating emergent information flows and discerning the actionable information contained within them. Each of the papers advances our understanding in this area. The team developed scripts to analyze the use of syntactical features and shorthand associated with small text communication like that found on twitter. Using the first set of scripts, they submitted an ACM Hyptertext Conference paper. A second set of scripts is more focused on URL decoding and information sharing by leaders using small text communication. Initiated development of a paper focused on characterizing 25 corpora of electronic trace data in the Goggins lab, ranging from APAN data to Twitter, Facebook, online learning, software engineering and others. The goals of this paper is meta level description of how to quickly and operationally assess trace data and systematically integrate data from disparate sources, pertaining to the same phenomena, into an analytical workflow.

Goggins continues to develop papers, research methods and technical approaches to make the process of making sense of trace data more systematic and transparent for consumers of such analysis. During the past year Goggins has run three Open Data Hackathons at the iConference, ACM CSCW 2014 and as part of Philly Tech Week 2013 – all with the aim of developing a data factory approach to increase replicable and transparent analysis of leadership from trace data.

Task 3: Operationalization of emergence of shared leadership in teams (PSU, OSU, and Aptima)

Ohio State University has been primarily focused on using an automated algorithm developed in collaboration with Aptima, Inc. in order to advance the conceptual underpinnings of the theory of macrocognition by comparing what is the same and different with the PSU theoretical framework. The automated algorithm was primarily developed from a dataset from prior ONR funding on undergraduate students doing a logistics task as an ad hoc team. It has also been used on a dataset of resident physician sign-outs in an intensive care unit at the end of a shift. The algorithm identifies leaders and detects verbalizations that indicate complex phenomena in macrocognition based upon comparisons of theoretical concepts with Penn State University. In particular, the algorithm identifies collaborative cross-checks, a form of error detection that employs "fresh eyes" on a situation by an incoming team member to uncover erroneous assumptions. In the context of our theoretical framework of macrocognition functions, collaborative cross-checks uncover erroneous sensemaking activities (e.g., patient diagnoses), inappropriate elements of treatment plans, particularly with respect to not taking into account time horizons when planning (e.g., placement to home without taking into account patient prognoses).

In this work, operationalization of leadership was conducted at four levels:

Information Transfer (IT)- *How new info existing prior to collaboration is added.*
(AI): Add Info- Add new information w/o prompting (PUSH ACTS)
(Q): Question- Prompt someone for new information
(R): Reply- Provide new information in response to a prompt (PULL ACTS)

2. Check Understanding (CU)- *How previously added info is checked or repaired.*
(CH): Check- verifying understanding*
(CL): Clarify- clarifying or restating information (AI info)*
(AC): Acknowledge- signaling receipt or understanding of information

3. Management of Processes (MP)- *How work is orchestrated*
(MN) Management- discussions centered on interactions or how to do the work
(CM) Command for action, order or instruction that does not take others into account
(RQ) Request for action- posed as a question or indirect prompt (not a question)

4. Interpretation & Decision Making (ID)- *How task information is interpreted*
(J): Judge- Individual preference, opinion, or claim, with or without deliberation
(RA): Rational that supports a judge (J) or alternative (AT) act.
(AT): Proposing alternative to a (J) OR (RA) act.
(CO): Confirmation- Requesting agreement on a proposed decision

(AG): Agreement- Indicating agreement for prior judgment or decision

A finding from both the OSU and PSU datasets using different domains with the leadership identification algorithm is that some leaders do meta-cognitive anchoring and timekeeping functions, whereas other leaders delegate these to another team member. It is anticipated that using the following categories would improve the automated algorithm based upon comparing the results from OSU's automated codes and PSU's manually generated codes for the following similar concepts in the codebooks:

- "Ask clarifying question" (OSU) and "Checking/clarifying information" (PSU)
- "Collaborative cross-checking" (OSU) with "Request evidence/management process" (PSU) and "Alternative theories/decision making activity " (PSU)
- "Off-topic" (OSU) with "Other" (PSU)
- "Identification" (OSU) and "Management of process" (PSU)

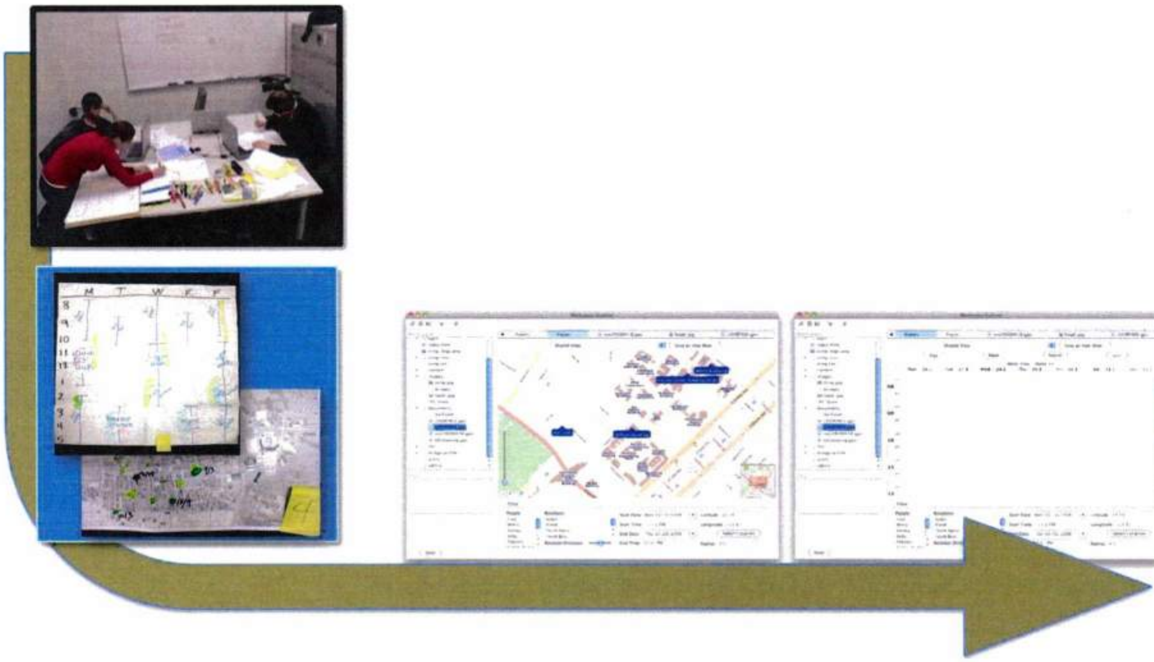


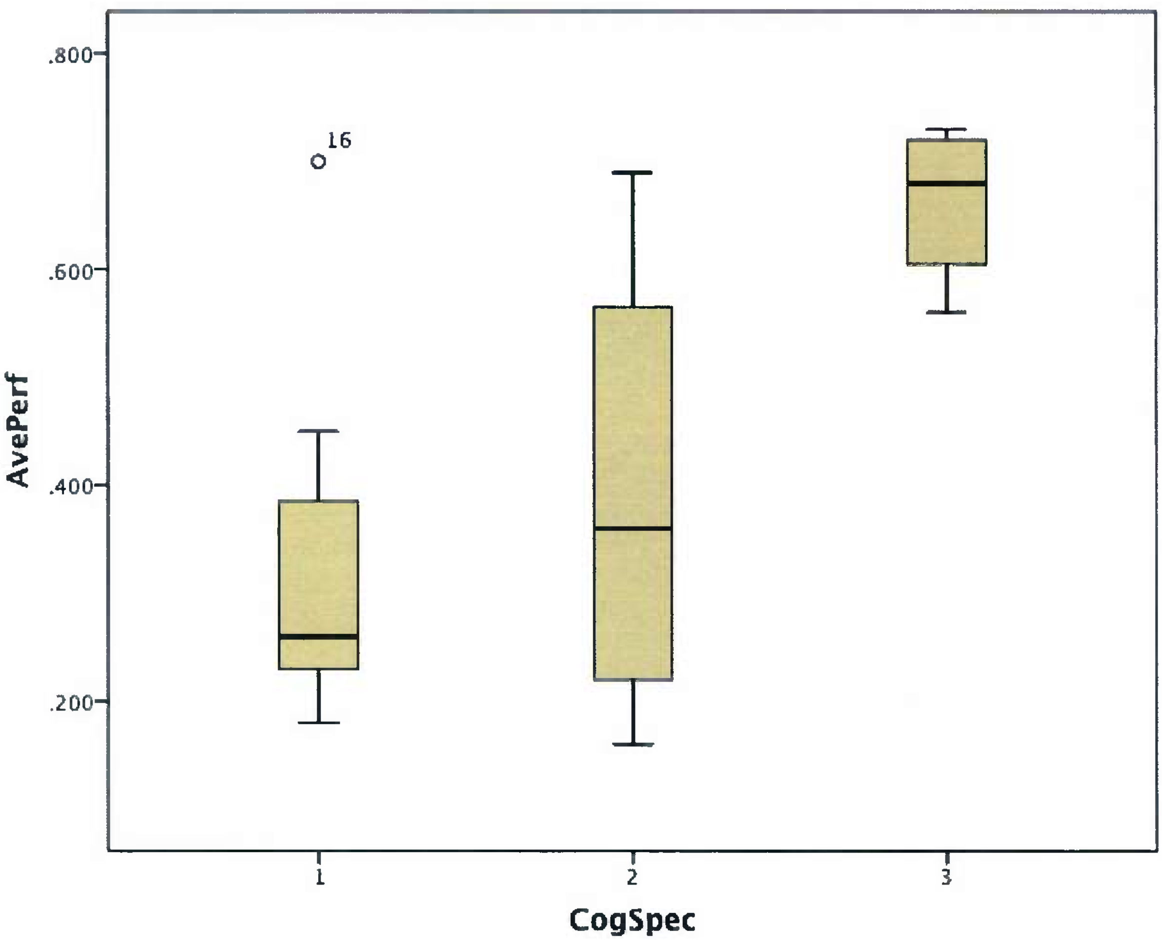
Figure 3 From operationalization of shared leadership to design of support at Penn State

In separate work, Co-PI Borge conducted a rigorous study examining complex collaborative decision-making under time pressure. Her team collected various sources of data with the aim to use observations of human behavior including conversational data as a means to pull out requirements for the design of cognitive support tools. They also made sure that the types of behaviors and processes exhibited by their participants coincided with real-world activity. They conducted a micro-analysis of 20 teams, which included over 70 hours of video, resulting in over 34,000 dialogue acts.

Penn State's findings suggest that cognitive specialization is a more critical variable than verbal equity. They found no significant differences between the verbal equity of high and low

performing teams, but did find significant differences in cognitive specialization. Complete cognitive specialization, where teammates controlled separate, complex cognitive activities was associated with their highest performing teams. Their highest performing teams had one member control cognitive activities associated with orchestration and executive control of task objectives, i.e., sociocognition and metacognition, and a different member control cognitive activities

associated with completing task objectives, i.e., cognitive behaviors. In contrast, our lowest performing teams had one member control both types of cognitive activity. Complete domination by one member was primarily associated low performance (see box plot figure below). Patterns indicating sharing of cognitive responsibilities, where one member controlled one activity but maintained active participation in another was associated with both high and low-performing teams.



A box plot comparing average performance and cognitive specialization. Only one of our highest performing teams, case 16, did not show cognitive specialization. Whereas most of the lower performing teams showed little to no cognitive specialization.

Findings regarding the range of sophistication of cognitive behaviors indicate that cognitive specialization (shown in the table below) may be associated with more sophisticated cognitive behaviors. This suggests that simply sharing cognitive responsibilities may not be helpful for a

List of the cognitive behaviors exhibited by two teams exhibiting similar input characteristics, but demonstrating the two extremes of cognitive specialization.. A column is also included that indicated whether the behavior was commonly seen across multiple teams.

Types of Cognition	Behavior	Definition	Team 21	Team 2	Common
Cognitive	Accretion (Carroll et al., 2013)	Act of recording: inscribing verbalized information unto the artifact "as is" without data reduction strategy. May continue to add more information or rules to artifact.	X	X	X
	Fact retrieval	To refer to a piece of shared information contained in the artifact as part of an information transfer or check/clarify behavior.	X	X	X
	Identify needed info	To use artifact as a means to deduce what other information pieces are necessary to search for.	X	X	X
	Support Claims	To pull specific information piece from artifact contents to use as rationale to support claim.	X	X	X
	Refute Claims	To pull specific information piece from artifact content to use as evidence against a claim.		X	
	Filtering/ Constraining interpretation (Ainsworth, 1999; Carroll et al., 2013)	Act of filtering: To use inherent properties in the artifact to organize and exclude information from or to another artifact.		X	X
	Extension (Ainsworth,1999)	(Task) To make a generalization about people, events, or claims, etc. based on aggregated information contained in the artifact.		X	
Sociocognitive	Confirm	To use content on artifact to ensure proper understanding of another's claim.	X	X	X
	Repair	To use content on artifact to identify & correct misunderstanding or missing information previously stated.	X	X	X
	Anchor Talk	To use information contained in the artifact to make people aware of narrowing the topic of discussion to a specific person, place, event, or location on the artifact.	X	X	X
	Organize Talk	To use content of artifact to organize the order in which information is shared or which topics are to be discussed.		X	
Metacognitive	Artifact Decomposition	(Artifact) To identify and organize aspects of the artifact, such as features, symbols, and color-coding.	X	X	X
	Monitor	To use artifact to Make a meta comment regarding amount of information shared, reliability of information identifying missing information, or what remains to be done.	X	X	
	Task Decomposition	(Task) To identify & organize variables of the task in the artifact as a means to break down the task into smaller ordered sub-tasks.		X	
Cognitive Event	Cognitive Linking (Kaput, 1989)	To use multiple representations (more than two) to link information across artifacts during decision making processes		X	

team. The goal should be to maximize cognitive power through specialization.

The complexity of the task also made it a good data source for Ohio and Aptima to test their models and compare their findings to the original manually coded findings. Their data served as a means for Duchon and Patterson to test their trained model on a different data set and separate task. The Borge data set was very rich in terms of including a wide variety of characteristics of individuals within teams, which made it possible to examine how patterns in interaction and

leadership taking were related to team composition in terms of combinations of these personal characteristics.

Type of Coding	People Required	Prep Time	Coding Time	Total Work Hours
Penn State: Manual Coding	5	3 Work Months	8 Months	3300 Hours
Aptima: Automated Coding	1	15 Work Days	Seconds	100-120 Hours

This table shows the extent to which the cost of doing communication analysis can be reduced by developing computational models to automate the annotation for you.

MN= Task Management (95% match) , DEC= Decision-Making Behavior (80% Match)

Overall Performance	Aptima (Automated selection)	Penn State (Manual Selection)
H	WEB	WEB (MN) & Rec (Dec)
H	REC	REC (MN + Dec), INT (Dec)
H	REC	WEB (MN) & REC(Dec)
H	REC	REC (MN + Dec)
H	REC	REC (MN + Dec) & INT (Dec)
H	REC	REC (MN + Dec)
H	WEB	WEB (MN + Dec)
H	REC	REC (MN) & INT (Dec)
H	Web	WEB (MN) & INT & REC (Dec)
H	REC	REC (MN + Dec)
L	INT	INT (MN + Dec)
L	WEB	Web (MN + Dec) & REC (Dec)
L	REC	REC (MN + Dec)
L	REC	REC (MN + Dec)
L	REC	INT (MN + Dec) & REC (MN)
L	REC	REC (MN + Dec)
L	WEB	WEB (MN + Dec)
L	REC	REC (MN + Dec)
L	WEB	WEB (MN + Dec)
L	REC	REC (MN + Dec)

Figure 4 Match between hand identified leaders (Borge hand analysis) and automatically identified leaders (Duchon and Patterson model)

For simply identifying the leader in a group, as displayed in Figure 4, the automated model they trained on Ohio state data worked EXTREMLY well for the Penn State data in identifying one

primary leader. If we can reliably identify who is making decisions and managing groups via textual information, we could identify leaders within an organization or cell. This can provide us with information about who holds decision-making power and who should be targeted for consultations or as a person of interest.

This model was applied by Aptima to the forum communications from APAN as well. In particular it was combined with a standard LDA topic model and aggregated by the type of organization of the sender (US Military, US Government, NGO, etc.). This analysis showed for example that the Military did provide the most leadership overall and on most topics, except one topic related to families and children, for which NGOs understandably evidenced more leadership.

Borge & Carroll continue to develop papers related to this work. They recently submitted a paper to the International Association for Computing Machinery Conference for supporting Group Work, 2014. Borge was also invited to chair a session at the American Educational Research Association for work related to improving metacomprehension strategies and discuss research findings related to this project. Borge has three additional papers in progress. Borge and Duchon are also collaborating on a paper where they examine the reliability and validity of automatic detection of high quality decision-making processes. One of the preliminary findings from the shared datasets includes both qualitative and quantitative evidence to support the claim that high performing teams can be automatically detected by examining the relationship between idea building and critical evaluation.

Task 4: Develop support for leadership taking and shared leadership in teams

Automatic detection of opportunities for shared leadership (Penn State, Aptima, and OSU)

The Borge dataset described above was a valuable comparison case with earlier analyses conducted by Duchon and Patterson on a data set collected at OSU because the interaction was at least partly face-to-face rather than over chat. Thus, in our application of the Duchon and Patterson model for automated identification of leaders, some key information was not usable by the model, which was designed for purely textual interaction. Dealing only with textual data means you lose visual cues, making leadership detection more difficult. Also since groups can share a lot of information – simply giving a report or sharing lots of information (i.e. number of contributions) does not entail leadership taking to the same extent that it might in a purely text based interaction context. Nevertheless, the automated analysis achieved a high level of accuracy, and demonstrates the feasibility of achieving efficient analysis automatically. An error analysis indicated some key limitations of the Duchon and Patterson model, which may be addressed by alternative computational modeling approaches. Using LightSIDE, we have the ability to experiment with a wide variety of modeling approaches in order to work towards higher performance and better transferability.

Borge and Duchon are in the process of examining patterns of idea building and idea negotiation in order to propose training requirements for collaborative problem solving teams. This work can inform the types of metacognitive and visual supports provided to new teams as a means to

develop their abilities to engage in higher quality problem solving activity. Intelligent training supports and automated feedback could minimize the likelihood of cognitive breakdowns at the team level and help to enhance team performance.

Intelligent agents for supporting macrocognition (Ohio State and CMU)

Last year, the CMU team had developed a more robust and easy to use version of the Basilica development framework used in their earlier work on using conversational agents to support group work (Adamson & Rosé, 2012). The new framework has been used in several recent successful studies of technology supported group work (Dyke et al., in press; Howley et al., 2012; Adamson et al., 2013; Adamson et al., 2014).

FINAL WORK PLAN/ FINAL PROJECT WRAP-UP

At the time of our last report, we had only a few months of funding left in the current grant. At that time, we reported the following plans, which we have followed through with. Currently we have no remaining funding.

Our plans for the last year and a half of our funding included finishing development of the data to actionable knowledge pipeline, further advances in machine learning and team science that will make the actionable knowledge derived from the pipeline more useful, and additional theoretical advances in analysis of emergence of leadership that will further improve our ability to develop technological support for effective teamwork. As we reported in the last report, we had made substantial progress on all of these fronts, and we worked to wrap up all of these things in our final months. In terms of dissemination, PI Rosé gave tutorials at the Learning Analytics Summer Institute and the new ACM Learning@Scale conference where she disseminated the framework for analysis of leadership in teams developed at CMU.

One of the most important and exciting next steps, which will be spearheaded by Co-PI Duchon is bringing our analytic engine into the context of a real-time Army exercise. We will also continue to collect corpora from community members and work on publications of cross-analyses already conducted as well as new ones we will continue to work on in our time remaining.

One important development goal has been to complete the data to actionable knowledge loop by enabling remote users who access and analyze datasets to import their analyses back into the Aptima database. We released LightSIDE 2.0, which newly included some of the most recent algorithms for domain adaptation and multi-domain learning built in. Drexel has prepared its Group Informatics Toolkit as an R (r-project.org) package in FY2013, and continued to work with Aptima on prototyping real time applications of the developed API. We have worked to iteratively improve the user experience with the shared analysis pipeline linking the Aptima database with analysis tools and will continue to do so. Specifically, this has and will continue to involve small scale (informal) user studies and refinement. We will continue to collect, process, and insert additional datasets from the CDM community.

One important focus for technical research related to machine learning has been to address not only generalization across sub-community structures within a hierarchical dataset, but also accommodating changes over time (in the form of evolving behavior models, not just how individual sub-populations evolve and change, but how they evolve and change in relation to and in response to one another) in a longitudinal dataset. For this work, we have been building on the results reported in (Jain et al., 2012) using latent variable modeling techniques. We have two prototype models built and working and are now extending that work to make it more scalable to larger datasets. In the final months of the project, we completed such a model, which is highly scalable, being able to be applied to networks with millions of users. We validated the model on 3 different datasets from Massive Open Online Courses and found that the subcommunity

structure identified by the algorithm was predictive of differences in dropout rate between subsets of students.

The technical infrastructure we have built up is already enabling transfer of codes from one dataset to another, and comparison between leadership constructs. Now that we have the capability to do this work, we can use these tools to advance understanding of the similarities and differences of constructs, and to begin to make progress towards an integrated, unified framework for the analysis of shared leadership. We reported results on application of a model trained on Emily's data to data from Marcela's group. This revealed some important limitations in the model that we are working on addressing.

Next steps for future work might include having Borge collaborate with Duchon and Patterson as well as Rosé to improve the current model. We have some hypotheses as to why we are seeing certain patterns and connections between our findings. Through collaboration between Aptima, OSU, and CMU the Penn State team could evaluate the model and see if it is possible to train the system to identify more than one leader or different types of leaders. This would allow us to make more accurate identifications. We could also leverage the Aptima database as a means to do the training. More interestingly, we might get better at reliably identifying judgments at the level of the utterance- if so, this could be used as a means to support collaborative interactions and cognition in general.

Goggins recently moved to a new tenure track position at the University of Missouri's Informatics Institute and School of Information Science and Learning Technologies. Since concluding work on this grant in September, 2013, Goggins is continuing to develop papers and technologies emerging from work with this team and is building models of the reflexive dynamics of teams. Specific efforts include the recent implementation of a topic modeling + network analysis algorithm in the R Statisical Software package, leading the organization of the Consortium for the Science of Sociotechnical Systems Summer Institute and Co-Chairing ACM Group 2014.

MAJOR PROBLEMS/ISSUES (250 WORDS)

No major problems since the last report.

TECHNOLOGY TRANSFER

Our biggest success with technology transfer is the inception of the LightSIDE labs startup company, which is focusing on developing enterprise software solutions that build on machine learning models developed using LightSIDE. LightSIDE has been downloaded over 5,000 times, and LightSIDE labs has large contracts with College Board and McGraw Hill, with others in progress. Borge has also been recruited by a learning technology company, CoLearnr. She will draw on findings from this study to inform the design of a collaborative information synthesis tool.

Beyond this, we would still be extremely interested in partnering with organizations like APAN that provide collaboration environments to house relief efforts so that we could tune our infrastructure to meeting real contextual needs in relief efforts.

We would also like to continue partnering with CDM researchers outside of our project group so that we can iteratively improve the support we are offering for data analysis and development of supportive interventions to house distributed collaboration. Looking to the future, we would like to extend our capabilities for supporting distributed collaboration within homogeneous online environments (i.e., where all team members have access to the same shared interface and comparable resources) to heterogeneous environments where some team members have access to rich resources in a full online environments, and others are working on the ground with limited connectivity, perhaps through a simple SMS-based connection. This raises interesting questions related to coordination of multiple distinct perspectives, where there is significant inconsistency between perspectives, and great uncertainty of information.

The database infrastructure and webservices being developed here also mean that techniques developed will be more easily transferred to new domains, both military and civilian. For example, through other projects, Aptima gathers data at one or two military exercise per year using this infrastructure to store communications. This means that any techniques which work with those webservices could also tap into those data, and provide their analyses to drive leadership identification, information routing, and other context-sensitive support.

Foreign Collaborations and Supported Foreign Nationals

The only foreign nationals supported on this grant are graduate students who have been involved in the work. Their names are indicated below under Award Participants.

Productivity

Books

- Suthers, D., Lund, K., Rosé, C. P., Teplovs, C., Law, N. (2013). *Productive Multivocality in the Analysis of Group Interactions,* edited volume, Springer.

Journal articles

- Goggins, S., Petacovic, E. (2014). Connecting Theory to Social Technology Platforms: A Framework For Measuring Influence in Context. *American Behavioral Scientist,* Accepted.
- McDonald, N., Blincoe, K., Goggins, S. (2014). Modeling Distributed Collaboration on GitHub. *Advances in Complex Systems,* Accepted.
- Welch SJ, Cheung DS, Apker J, Patterson ES. (2013). Strategies for Improving Communication in the Emergency Department: Mediums and Messages in a Noisy Environment. The Joint Commission Journal on Quality and Patient Safety. Vol. 39, no. 6. 279-286.

- Wu, A., Convertino, G, Ganoe, C.H., Carroll, J.M. & Zhang, X. 2013. Supporting collaborative sensemaking in emergency management through geo-visualization. International Journal of Human-Computer Studies, Special Issue on Shared Representations, 71 (1), 4-23.
- Patterson ES, Hoffman R. (2012) Modeling Macrocognitive Funtions of Distributed Cognition to Navigate Adaptations to Change and Uncertainty, Cognition, Technology and Work. Vol. 14, no. 3: 221-227.
- Mascaro, C., & Goggins, S. (2013). Coffee or Tea: The Emergence of Networks of Discourse in Two Online Political Groups. Journal of Information Technology and Politics, accepted
- Kumar, R. & Rosé, C. P. (accepted). Triggering Effective Social Support for Online Groups. *ACM Transactions on Interactive Intelligent Systems.*
- Goggins, S., Valetto, P., Mascaro, C., and Blincoe, K. (in press). Creating A model of the Dynamics of Socio-Technical Groups using Electronic Trace Data. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research.*
- Gweon, G., Jain, M., Mc Donough, J., Raj, B., Rosé, C. P. (2013). Measuring Prevalence of Other-Oriented Transactive Contributions Using an Automated Measure of Speech Style Accommodation, *International Journal of Computer Supported Collaborative Learning* 8(2), pp 245-265.
- Carroll, Borge, & Shih (2013). Cognitive Artifacts as a Window on Design. Journal of Visual Languages and Computing, http://dx.doi.org/10.1016/j.jvlc.2013.05.001i
- Goggins, S., & Jahnke, I. (2012). CSCL@ Work: Making Learning Visible in Unexpected Online Places Across Established Boundaries. International Journal of Sociotechnology and Knowledge Development (IJSKD), 4(3), 17–37.
- Anders S, Schweikhart S, Woods DD, Ebright P, Patterson ES. (2012) The Effects of Health Information Technology Change Over Time: A Study of Tele-ICU Functions." Applied Clinical Informatics. Vol. 3, 239-247.
- Mayfield, E., Laws, B., Wilson, I., & Rosé, C. P. (2014). Automating Annotation of Information Flow for Analysis of Clinical Conversation, *Journal of the American Medical Informatics Association 21 (1), pp 122-128.*
- Adamson, D., Dyke, G., Jang, H. J., Rosé, C. P. (2014). Towards an Agile Approach to Adapting Dynamic Collaboration Support to Student Needs, *International Journal of AI in Education* 24(1), pp91-121.

Non-refereed significant publications

- None

Book Chapters

- Rosé, C. P. & Tovares, A. (in press). What Sociolinguistics and Machine Learning Have to Say to One Another about Interaction Analysis, in Resnick, L., Asterhan, C., Clarke, S. (Eds.) *Socializing Intelligence Through Academic Talk and Dialogue,* Washington, DC: American Educational Research Association.

- Rosé, C. P. (in press). A Multivocal Analysis of the Emergence of Leadership in Chemistry Study Groups, in Suthers, D., Lund, K., Rosé, C. P., Teplovs, C., Law, N. (Eds.). *Productive Multivocality in the Analysis of Group Interactions*, edited volume, Springer.
- Mayfield, E. & Rosé, C. P. (2013). LightSIDE: Open Source Machine Learning for Text Accessible to Non-Experts, Invited chapter in *the Handbook of Automated Essay Grading.*
- Rosé, C. P. & Lund, K. (2013). Methods for Multivocality, in Suthers, D., Lund, K., Rosé, C. P., Teplovs, C., Law, N. (Eds.). *Productive Multivocality in the Analysis of Group Interactions,* edited volume, Springer.
- Lund, K., Rosé, C. P., Suthers, D., & Baker, M. (2013). Theoretical perspectives on multivocality, in Suthers, D., Lund, K., Rosé, C. P., Teplovs, C., Law, N. (Eds.). *Productive Multivocality in the Analysis of Group Interactions,* edited volume, Springer.
-

Technical Reports

- None

Workshop and Conference Papers

- Borge, M., Goggins, S. (Accepted). Developing a Community of Learners With Social Media. Submitted to *The 11th International Conference of the Learning Sciences.*
- Graves, I., McDonald, N., & Goggins, S. (2014). Sifting signal from noise: a new perspective on the meaning of tweets about the 'big game'. *New Media & Society, Accepted.*
- Black, A., Mascaro, C., Gallagher, M., and Goggins, S. (2012). Twitter Zombie: Architecture for Capturing, Socially Transforming and Analyzing the Twittersphere. *ACM Group 2012.*
- Duchon, A., and Patterson, E. S. (2014). Identifying Emergent Thought Leaders. In: W.G. Kennedy, N. Agarwal, and S.J. Yang (eds.) *International Social Computing, Behavioral Modeling, and Prediction Conference (SBP). Lecture Notes in Computer Science* 8393, 50-57.
- Goggins, SP. (2013). Collaboration in Isolation: Bridging Social and Geographical Boundaries in Two Rural Technology Firms. 2013 iConference
- Goggins, S., Mascaro, C., and Mascaro, S. (2012). Relief after the 2010 Haiti Earthquake: Participation and Leadership in an Online Resource Coordination Network. *ACM Conference on Computer Supported Cooperative Work.* 57-66.
- Jain, M., McDonogh, J., Gweon, G., Raj, B., Rosé, C. P. (2012). An Unsupervised Dynamic Bayesian Network Approach to Measuring Speech Style Accommodation, in the *Proceedings of the European Association for Computational Linguistics*
- Joshi, M., Dredze, M., Cohen, W. & Rosé, C. P. (2012). Multi-Domain Learning: When Do Domains Matter, in *Proceedings of EMNLP: Conference on Empirical Methods in Natural Language Processing and Natural Language Learning*
- Joshi, M., Dredze, M., Cohen, W. & Rosé, C. P. (2013). What's in a Domain? Multi-Domain Learning for Multi-Attribute Data. *Proceedings of the North American Chapter of the Association for Computational Linguistics*

- McDonald, N., & Goggins, S. (2013). Performance and participation in open source software on GitHub. In CHI '13 Extended Abstracts on Human Factors in Computing Systems (pp. 139–144). New York, NY, USA: ACM. doi:10.1145/2468356.2468382
- Patterson, ES, Bernal F, Stephens R. (2012) Differences in Macrocognition Strategies With Face to Face and Distributed Teams, in Proceedings of the Human Factors and Ergonomics Society, 282-286.
- Patterson ES, Rayo MF, Weiss C, Woods Z, Mount-Campbell AF. Online Training for Resilience Communication Strategies during Shift Change Handovers, in Proceedings of the Human Factors and Ergonomics Society Annual Meeting. (Sep 2013). 57 (1). 1427-1431.
- Woods Z, Beecroft N, Duchon A, Hilligoss B, Patterson ES. Automatically detecting differences in communication during two types of patient handovers: A linguistic construct categorization approach. (in review for Human Factors and Ergonomics Society conference, submitted March 2014).
- Piergallini, M., Gadde, P., Dogruoz, S., Rosé, C. P. (2014). Modeling the Use of Graffiti Style Features to Signal Social Relations within a Multi-Domain Learning Paradigm, *Proceedings of the European Chapter of the Association for Computational Linguistics*
- Adamson, D., Bharadwaj, A., Singh, A., Ashe, C., Yaron, D., Rosé, C. P. (2014). Predicting Student Learning from Conversational Cues, *Proceedings of Intelligent Tutoring Systems*
- Yang, D., Wen, M., Rosé, C. P. (2014). Peer Influence on Attrition in Massively Open Online Courses, *Proceedings of Educational Data Mining*
- Mayfield, E., Adamson, D., & Rosé, C. P. (2013). Recognizing Rare Social Phenomena in Conversation: Empowerment Detection in Support Group Chatrooms, *Proceedings of the Association for Computational Linguistics*

Patents

- None

Awards

- Emerald Outstanding Paper Award (Goggins and colleagues)
- Rosé's team (with LightSIDE analysis tool developed under this grant) was invited participant as the sole non-commercial vendor in a nation wide automated essay grading grand challenge (news coverage in the National Public Radio and Education Week)
- Rosé's team (with LightSIDE analysis tool developed under this grant) started a company and tied for second best university based startup company at the Three Rivers Venture Faire.
- John M. Carroll was awarded the title of Distinguished Professor of Information Sciences and Technology.
- Emily S. Patterson was awarded the 2013 Faculty Scholarly Activity Award from the School of Health and Rehabilitation Sciences, Ohio State University College of Medicine.

Press Coverage and Other Publicity

- Rosé, Interactive TV appearance: Interviews on Gates Foundation funded interactive TV series produced by In the Telling: "Massive and Open: What are we learning?", part of a larger series aired on Internet TV called e-literate TV (filmed in December 2013).

- Rosé, Press Coverage: Profile Piece published in The New Learning Times, November 2013.

In preparation or Submitted articles

- Borge, M. (in preparation). Computer supported collaborative environments: Implications for future research designs. To be submitted to the International Journal of Computer Supported Collaborative Learning.
- Borge, M., and White, B.Y (under review). Sociocognitive Managerial Roles: An Approach to Developing Collaborative Competence. Submitted to *Cognition and Instruction.*
- Borge, M., & Carroll, J.M. (under review). Verbal Equity, Cognitive Specialization and Performance. Submitted to the ACM Group 2014 Conference.
- Borge, M., Duchon, A., & Carroll, J. *(in preparation). Automated* Identification of emergent leaders. To be submitted to the ACM SigCHI 2014 Conference.
- Goggins, S.P. (Under Review). Leadership Patterns Across Corpora: Toward a Meta Analytic Approach to Analysis of Socio-Technical Behavior Across Platforms. The Journal of User Modeling and Personalization.
- Goggins, S.P., & . (Under Review). Connecting Theory to Social Technology Platforms: A Framework For Measuring Influence in Context. American Behavioral Scientist, Under Review.
- Goggins, S.P., McDonald, N.K., & Valetto, G. (Under Review). Structural Fluidity in Virtual Organizations: A Case from Github. In CSCW 2014. Presented at the CSCW 2014, Baltimore, D.
- Mascaro, C, McDonald, N.K., & Goggins, S.P. (Under Review). What the Hashtag: Examining Hashtag Position on Twitter. Presented at the Hawaii International Conference on System Science, Hawaii: IEEE.
- McDonald, N.K., Blincoe, K., & Goggins, S. P. (Under Review). Modeling Distributed Collaboration on Github. Advances in Complex Systems.
- McDonald, N.K., & Goggins, S.P. (Under Review). Pull Requests and Participation in Github: Manifestations of Leadership in Open Source Software. In CSCW 2014. Presented at the CSCW, Baltimore: ACM.
- Rosé, C. P. & Borge, M. (in preparation). Invited chapter in E. Salas & Fiore, S. (Eds.) Measuring Engagement in Social Processes that Support Shared Cognition, *Developing Multidisciplinary Measurement Approaches for Team Cognition Research*, American Psychological Society.

Presentations (other than papers)

- Borge, M. Invited to chair a session on: Strategies to improve metacognition. Presented at the American Educational Research association, Philadelphia, PA, April 4-7th, 2014.
- Borge, M. Invited to present at a special session on gender equity: Stealth instruction

through games: WAGES (Workshop Activity for Gender Equity Simulation) Demonstrates Gender Inequity in the Workplace. To be presented at the 122nd annual convention of the American Psychological Association, Washington D.C.

- Borge, M. Invited discussant for the Waterbury Summit: Systems Thinking. Waterbury Summit. August 7-10, 2013, Pennsylvania State University.
- Borge, M. Invited talk on "Designing for Learning in Computer Supported Collaborative Environments". Presented to the College of Education, Pennsylvania State University as part of the Learning Sciences Group Speakers Series, April 2nd, 2012.
- Carroll, J.M. 2012. Humanity, technology and HCI. Honoris Causa Address, Universidad Carlos III de Madrid (September 18).
- Carroll, J.M. 2012. Activity Awareness. EnRiCH International Network for Collaborative Practice and Community Engagement Workshop (Ottawa, Canada, November 27-30).
- Carroll, J.M. 2013. The future of work. Keynote for 16th Congress of the European Association of Work and Organizational Psychology (EAWOP) to be held 2013 May 22nd-25th in Muenster, Germany
- Carroll, J.M., Hoffman, B., Robinson, H. & Han, K. & Rosson, M.B. 2013. Hyperlocality and Suprathresholding in Community Network Designs. CHI 2013 Workshop on Human Computer Interaction for Third Places (HCI3P 2013), Paris, France, April 27-28.
- Duchon, A. Invited Panelist on The Digital Frontier: Facilitating Teamwork through Bits and Bytes. Society for Industrial-Organizational Psychology Annual Conference. April 2013, Houston, TX.
- Duchon, A., Ganberg, G., Therrien, M. and Sullivan, S. C4: An Interoperable Communications Database for Sharing Data and Analyses. Presentation at the Interdisciplinary Network for Group Research Conference, Atlanta, July 2013.
- Goggins, S. (2014) "Panel: Crowdsourcing Crisis Response: The Boston Marathon Bombing" ICSCRAM 2014, State College, PA, May 20, 2014.
- Goggins, S. (2014) "Panel: The Ethos and Pragmatics of Data Sharing", ACM CSCW Conference, Baltimore, MD, Wednesday, February 19, 2014.
- Goggins, S. (2014) "Structural Fluidity and Performance in Virtual Organizations: Contrasting (and finding commonality) Between Disaster Scenarios and Open Source Software Projects", University of Indiana, Bloomington, IN, February 14, 2014.
- Goggins, S. (2013) "Structural Fluidity and Performance in Virtual Software Organizations", University of Nebraska, Omaha, November 1, 2013.
- Goggins, S. Invited Panel Talk on Computational Social Science in the iSchools, February 2013, Dallas, Tx
- Goggins, S. Invited Talk on Big Social Data and Computational Social Science, University of Missouri, January 2013
- Goggins, S. Invited Talk on Leadership Detection from Open Source Repositories and Social Media, University of Washington, Seattle, WA, May 2013
- Goggins, S. Invited Talk on Information Science, Libraries and Meta Data across Social Technology Platforms, University of Wisconsin, Madison, WI, June 2013
- Goggins S, Mascaro C, McDonald N, Black A, Valetto G. Big Social Data for Social and Information Scientists. Dallas, Texas: Illinois Digital Environment for Access to Learning and Scholarship; 2013.Working Group on "Big Social Data" organized, using the "bigsocialdata.org" web address

- Haynes, S.H., Carroll, J.M. & Mudgett, D. 2013. Evaluation Criteria for Safe Improvisation in EMS Technologies. CHI 2013 Workshop on Evaluation Methods for Creativity Support Environments (ECSE 2013), Paris, France, April 28.
- Patterson ES, Invited Presentation and Roundtable Discussion, Electronic health records and patient safety, Topical Research Symposium Healthcare Ergonomics and Safety, sponsored by NIOSH, May 8 2013, Cincinnati, OH
- Rosé, C. P. Invited Panel Talk, Panel on Translating collaborative project-based learning to online and blended environments, Workshop on Multidisciplinary Research for Online Education (MunROE, http://www.cra.org/ccc/mroe), sponsored by the Computing Community Consortium, Feb 11-12, 2013, Washington, DC
- Rosé, C. P. Invited Tutorial on Discourse Analytics, Learning Analytics Summer Institute (Co-Organized by the Society for Learning Analytic Research and Stanford University), July 2013, Stanford University
- Rosé, C. P. Invited Symposium Talk, Automated Approaches to Analyzing Data from Collaborative Learning Settings, Symposium on Trends in Support and Analysis of Collaborative Learning, Jointly organized by the Special Interest Groups on Instructional Design and Learning and Instruction with Computers, at the Biennial Meeting of the European Association for Research on Learning and Instruction, August 2013
- Rosé, C. P. Invited Workshop Talk, Measuring Engagement in Social Processes that Support Shared Cognition, Workshop on Developing Multi-Disciplinary Measurement Approaches for Shared Cognition, University of Central Florida, February 2013
- Rosé, C. P. Invited Instructor, Discourse Analytics: Assessment of Collaborative Learning Discussions, 2013 Academy of the German Institute for International Education Research, Salzschlirf, Germany, June 2013
- Rosé, C. P. Invited Feedback Panel Talk, Invited Workshop: How will Collaborative Problem Solving be assessed at international scale?, Workshop at the Computer Supported Collaborative Learning conference, June 2013
- Rosé, C. P. Invited Panel Talk, Zooming In and Out of Collaborative Process Analysis through Linguistically Informed Machine Learning Models, Invited Plenary Panel: To see the world and a grain of sand: Multiple methods in CSCL research, Computer Supported Collaborative Learning conference, June 2013
- Rosé, Keynote talk, Intelligent Tutoring Systems 2014, June 2014
- Rosé, Linguistically Informed Automated Analysis of Collaborative Learning Processes, Distinguished Lecture in the Software and Information Systems Department at UNC Charlotte, April 2014
- Rosé, Invited Talk/Visit, Lytics Lab, School of Education, Stanford University, March 13, 2014
- Rosé, Invited Talk, Human-Technology Partnership in Facilitation of Discursive Instruction, 2014 Cyberlearning Summit, June 2014
- Rosé, Invited Talk, School of Education, University of California at Irvine, March 14, 2014
- Rosé, Learning through Discussion: Foundations, Findings, and Future, Tutorial at the First Annual ACM Conference on Learning @ Scale, March 2014
- Rosé, Invited Talk/Visit at Educational Testing Service, invited by Alina von Davier, February 21, 2014
- Rosé, Invited Participant and presenter at the MOOC Workshop: Defining and Advancing Change (December 2013), with financial support from the Bill and Melinda Gates Foundation

- Rosé, C. P. Workshop Invited Talk, LightSIDE: Open Source Machine Learning for Text Accessible to Non-Experts, National Council on Measurements in Education Conference, Spring 2012, *talk delivered by Elijah Mayfield*
- Rosé, C. P. Workshop Invited Talk, Analysis of Social Positioning in Interaction, Indo-US Workshop on Large Scale Data Analytics and Intelligent Services, IISc, Bangalore, Dec 18-20, 2011
- Rosé, C. P. Invited paper presentation, What Sociolinguistics and Machine Learning Have to Say to One Another about Interaction Analysis, Socializing Intelligence Through Academic Talk and Dialogue Conference, sponsored by the American Education Research Association, September 2011

AWARD PARTICIPANTS (PLEASE LIST ALL UNDERGRAD AND GRAD STUDENTS, FACULTY, AND STAFF RECEIVING FINANCIAL SUPPORT FROM THIS PROJECT)

CMU
- Carolyn P. Rosé (PI)
- Mahesh Joshi (PhD student)
- Abhimanyu Kumar (MS student)

Penn State
- Marcela Borge (PI)
- John Carroll (PI)
- Scott Cunningham (Undergraduate Student)
- Anthony Sanchez (Undergraduate Student)
- Sean Thompson (Undergraduate Student)
- Todd Shimoda (Professional Programmer)

Aptima
- Andrew Duchon (PI)
- Mike Therrien (staff)
- Jeff Rousseau (staff)
- Diane Miller (staff)
- Matt Jacobsen (staff)

Drexel
- Sean Goggins (PI)
- Nora K. McDonald, PhD Candidate
- Christopher Mascaro, PhD Candidate
- Michael Gallagher, MS Recipient
- Alan Black, PhD Student

Ohio State University
- Emily S. Patterson (PI)